✓ **Latest release**

Australian Standard Classification of Languages (ASCL)

This classification is used for the collection, storage and dissemination of statistical and administrative data on languages spoken in Australia

Reference period: 2016

Released 18/07/2016

Australian Standard Classification of Languages (ASCL) – Major Review

Commencing in 2023, the ABS will undertake a major review of the Australian Standard Classification of Languages (ASCL). The review aims to update the ASCL to better reflect languages that are widely used in the Australian community.

Why does the ASCL need to be updated?

The ASCL was first published in 1997, with minor reviews conducted between 2005 and 2016. During this time, the Australian community has changed considerably and the ASCL needs to be updated to reflect these changes. The ABS has also received feedback from individuals and organisations identifying areas that need to be revised. Data from the 2021 Census will further inform the review.

Next steps

The first round of public consultation is scheduled to commence in September 2023. At this time, the ABS will invite feedback through the ABS Consultation Hub, regarding the high priority areas (scope) of the ASCL. More information will be available at Consult.abs.gov.au (https://consult.abs.gov.au/) in the coming weeks, including information about how to participate in the review process. There will be another consultation process when the proposed changes are released in mid-2024.

If you would like more information, please email the ABS at standards@abs.gov.au (mailto:standards@abs.gov.au) .

Overview

About the classification

The first edition of the Australian Standard Classification of Languages (ASCL) (ABS Cat. No. 1267.0) was published in 1997 to meet a statistical and administrative need for a classification of languages. It was designed for use in the collection, aggregation and dissemination of data relating to language usage in Australia and to classify the following language variable:

- First Language Spoken
- Languages Spoken at Home
- Main Language Spoken and
- Main Language Other than English Spoken at Home

ASCL is used within the ABS and by other organisations in the fields of health, community services, and education in a number of administrative and service delivery collections.

The classification was developed through extensive research, stakeholder consultation and data analysis including the use of Census of Population and Housing data to separately identify languages used in Australia by a significant number of people. When ASCL was first published, it was recognised that it would be necessary to add languages to the classification as Australia's migration patterns changed. Consequently, ASCL was revised in 2005 and 2011.

Examination of the 2011 Census of Population and Housing data, information from stakeholders and external sources indicated that some aspects of the classification required changes to improve its usefulness. As a result, a minor review of ASCL was undertaken in 2015-16.

Definition of language

While the ASCL does not attempt to offer an exhaustive definition of language, the following definition encompasses the essential elements of language as used in ASCL.

The Macquarie Dictionary (Sixth Edition, 2013) defines language as:

"Communication in the distinctively human manner, using a system of arbitrary symbols with conventionally assigned meanings, as by voice, writing, or sign language. Any set or system of such symbols as used in a more or less uniform fashion by a number of people, who are thus enabled to communicate intelligibly with one another."

The term "language" is used to describe the base (finest) level categories in ASCL. They include:

- those languages which are universally recognised as distinct and separate languages
- separately identified dialects
- creoles and pidgins
- groups of linguistically distinct languages (e.g. American languages)
- sign languages.

A dialect is a regional or social variety of a language distinguished by pronunciation, grammar, and/or vocabulary.

Some language entities considered dialects are listed as separate categories in ASCL for the following reasons:

- failure to separately include language dialects may decrease the usefulness of language data by limiting analysis to the parent language only when a more detailed breakdown might be desired or required
- the boundary between a language and its dialects is not always clear or agreed, and
- the majority of stakeholders consulted preferred to include certain dialects as separate categories.

Scope of the classification

All world languages are in scope of the classification and languages with significant numbers of speakers in Australia are separately identified within the classification structure. Special attention has been given to separately identifying Australian Indigenous languages. Languages which are not separately identified are included in the most appropriate residual category

of the classification.

Extinct or dead languages spoken for religious or academic purposes are included in the most appropriate residual category of the classification. However, if sufficient numbers of an extinct or dead language are identified as spoken in Australia, it is separately identified in the classification, for example Latin.

Sign languages are defined as a communication system using gestures rather than speech or writing (The Macquarie Dictionary (Sixth Edition, 2013)), and are included in the classification.

Languages not commonly used as a means of general communication between people, such as computer languages, are excluded from ASCL.

About the review

Purpose of the review

A minor review of the Australian Standard Classification of Languages (ASCL) was undertaken to:

- separately identify emerging languages, based on changes in immigration patterns in Australia
- improve the coding index
- rename a number of categories to make the names more accurate
- identify diminishing languages in Australia
- improve the profile and coverage of Australian Indigenous languages based on stakeholder submissions, previous Census of Population and Housing responses and information about Indigenous language revival projects.

The review was an update only; there was no attempt to review the conceptual model underpinning the classification or to make major structural changes.

How it was done

The following research activities were undertaken when reviewing ASCL 2011.

Statistical analysis

Analysis of the aggregated responses to the 2011 Census of Population and Housing language question was undertaken. Languages within 'not elsewhere classified' categories which clearly recorded over 100 responses were added as new categories. Some exceptions were made if a language was of specific interest, such as Australian Indigenous languages, where the criterion for separate identification is three or more speakers.

Research

External research was conducted to:

- confirm the appropriate terminology to be used for categories in the classification and to assess the accuracy of the coding of languages at the broad and narrow group and language levels
- assess the accuracy of code assignments and linkages
- identify Australian Indigenous languages.

As a part of the Federal Government's approach to Closing the Gap, Australian Indigenous languages are supported through the Indigenous Languages and Arts (ILA) program. This program assists the revival and maintenance of Australian Indigenous languages by supporting community based language projects and resources. Australian Indigenous languages were investigated through the ILA program, queries from users of ASCL, and the online Aboriginal and Torres Strait Islander Languages database. Languages data and speaker numbers found in these sources were compared with information gathered through general research of Government, Australian Indigenous language, interpreter and academic sources.

Issues relating to non-Indigenous languages were identified from queries and submissions. Language data, including alternative spellings were investigated on the 'Ethnologue' database and other external web sites.

Stakeholder input

Relevant agencies, peak bodies and individuals with language expertise who had participated in the 2011 review were invited to make submissions to this review. In addition, the review was publicised on the ABS website inviting submissions from any interested party.

Consideration of suggestions received was limited to:

- speaker numbers
- new languages spoken in Australia
- growth and decline of languages, including Australian Indigenous languages
- index coverage
- alternative language names and spellings.

Submissions were analysed and reviewed and, where necessary, appropriate changes to the classification were made.

What has changed

Summary of changes

Being a minor review, no changes to the broad levels of the Australian Standard Classification of Languages (ASCL) were considered. Changes considered were limited to adding and removing languages, amending the names of some languages and adding appropriate entries to the coding index. These changes were based on Census of Population and Housing 2011 data, research from external sources, and stakeholder queries and suggestions.

Languages added to the classification

Australian Indigenous languages

One Australian Indigenous language has been added to the classification, namely:

- Yugambeh (8965) to Narrow Group 89 Other Australian Indigenous Languages.

Non-Indigenous languages

Two non-Indigenous languages have been added to ASCL:

- Zomi (6105) has been added to Narrow Group 61 Burmese and Related Languages
- Lingala (9262) has been added to Narrow Group 92 African Languages.

Languages which have been renamed

Indigenous languages

No Indigenous languages were renamed or re-described.

Non-Indigenous languages

To better reflect the languages in each group, the following languages have been renamed, based on research and stakeholder recommendations:

- Pitcairnese (9404) has been renamed Norf'k-Pitcairn
- Makaton (9702) has been renamed Key Word Sign Australia.

Changes to the coding index

Australian Indigenous languages

A number of changes have been made to the coding index relating to Australian Indigenous languages which include:

- the addition of alternative names and spellings for existing languages
- changes to the code assigned to index entries relating to Yugambeh (including alternate names and dialects) to reflect the

new language category

- changes to the code assigned to some index entries to correct errors and inconsistencies in coding some languages and dialects.

Non-Indigenous languages

A number of changes have been made to non-Indigenous languages in the coding index. These changes include:

- the addition of a large number of alternative names, name variants, and common misspellings of existing languages
- changes to the code assigned to index entries "Tai Shan" and "Taishanese" to correctly reflect these as a variant of 7101 Cantonese, rather than 6499 Tai nec
- changes to the code assigned to index entry "Kakwa" to correctly reflect this as a dialect of 9242 Bari, rather than 9299 African Languages nec
- changes to the code assigned to index entries relating to the new languages added to ASCL (including alternate names and dialects) to reflect the new language categories
- changes to the code assigned to a number of ambiguous responses to more appropriately assign a 'not further defined' code.

Comparing current and previous editions of ASCL

The ABS urges users and providers of language data to collect, classify and disseminate data using ASCL 2016 as soon as practicable/possible. There will be circumstances where users need to convert data from earlier versions of ASCL to ASCL 2016. To facilitate this process, a correspondence table between the 2016 and 2011 editions of ASCL is provided. In almost all cases, the languages of the two editions of the classification retain a one-to-one relationship. The correspondence table itemises the code linkages between the languages and indicates the movement of particular languages between groups in the two structures. The correspondence table is provided in the ASCL Data Cube, accessible from the Data downloads section.

In some instances, there is not a direct relationship between the language categories in the two editions of ASCL. Partial linkages at the language level are indicated by including the word 'part' after the code of the language concerned.

Building the classification

Minimum number of speakers threshold

The Australian Standard Classification of Languages (ASCL) does not list all (or even most) of the approximately 6,000 languages spoken worldwide. In order to be separately identified in ASCL, a non-Indigenous language must have 100 or more speakers in Australia. For Australian Indigenous languages the minimum threshold is three known speakers.

Classification structure

The ASCL has a three-level hierarchical structure, as follows:

Broad groups (one-digit codes)

The first and most general level of the classification comprises nine broad groups of languages, including one 'other' category. Broad groups are formed by aggregating geographically proximate narrow groups.

Narrow groups (two-digit and three-digit codes)

The second level of the classification comprises 51 narrow groups of languages similar in terms of the classification criteria, including seven 'other' categories which consist of languages which do not fit into a particular narrow group.

As there are a large number of Australian Indigenous languages, three narrow groups have been subdivided by inserting three-digit categories. The narrow groups affected are:

- 81 Arnhem Land and Daly River Region Languages
- 82 Yolngu Matha and

- 86 Arandic

There are 13 such three-digit categories in total and they provide meaningful and useful groupings within these narrow groups as well as allowing greater flexibility in presenting or summarising data for these Australian Indigenous languages.

Languages (four-digit codes)

The third and most detailed level of the classification consists of 435 languages, including 44 'not elsewhere classified' (nec) categories. The 435 languages include:

- 217 Australian Indigenous languages (including 20 nec categories), and
- 218 non-Indigenous languages (including 24 nec categories).

This is an increase of three languages since the 2011 edition and includes one additional Australian Indigenous language.

A pictorial representation of the ASCL structure, including examples, is shown below:
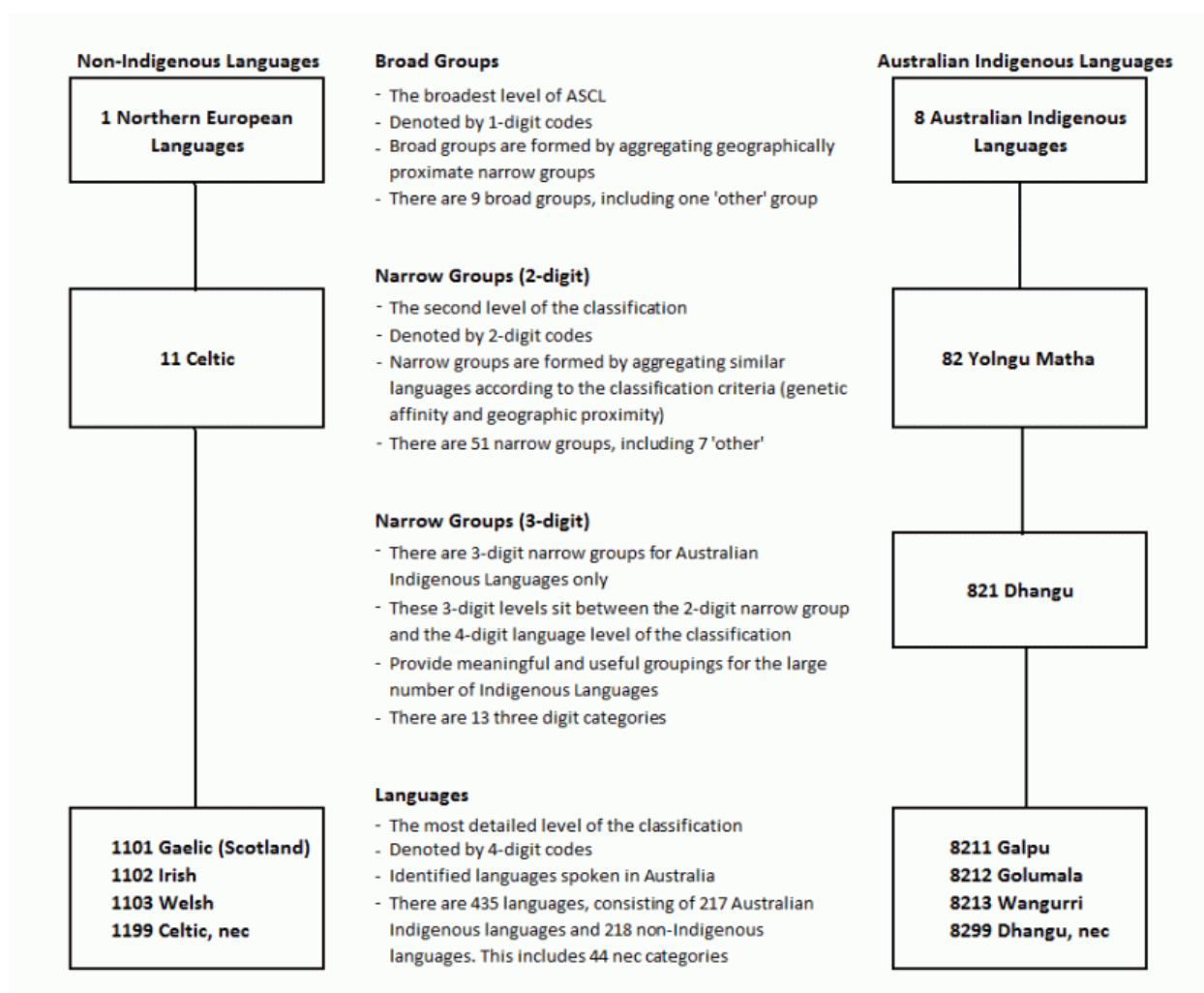
The full classification is available in the ASCL data cube, accessible from the Data downloads section.

Classification criteria and their application

Classification criteria are the principles by which categories are aggregated to form broader categories within a classification. The classification criteria used in ASCL are:

- the relationship between languages as a result of their evolution from a common ancestral language (genetic affinity)
- the area in which a language originated (geographic proximity). This also refers to the area where a language was first acknowledged as a distinct entity.

In the ASCL, languages are grouped into progressively broader categories, generally on the basis of genetic affinity and the geographic proximity of areas where particular languages originated. This allows populations of language speakers whose languages have evolved from common linguistic roots to be grouped in analytically useful ways. Secondary use of geography at the narrow group level also enables the formation of more meaningful residual language categories.

For usability purposes in the Australian context, the classification criteria have not been applied strictly in Broad Group 9 - Other Languages (see Residual Categories).

Residual categories

Broad Group 9 Other Languages consists of groups of languages which are not linguistically or geographically related and do not have sufficient speakers in Australia to form separate broad groups. At the narrow group level, the residual categories contain languages which originated in the same geographic area but which are linguistically unrelated to other languages in the broad group.

At the language level of the classification, the residual categories comprise languages which are genetically related and geographically proximate to the other languages in the narrow group. However, these languages have not been separately identified in the classification because they do not have sufficient numbers of speakers in Australia to form a category of their own.

About codes

Code scheme

The Australian Standard Classification of Languages (ASCL) coding scheme, with examples, is as follows:

Non-Indigenous languages

Code length	Hierarchical level	Example
1-digit	Broad group	1 Northern European Languages
2-digit	Narrow group	11 Celtic
4-digit	Language	1101 Gaelic (Scotland)

Australian Indigenous languages

To accommodate the large number of Australian Indigenous languages, and to facilitate greater analysis, three narrow groups within broad group 8 Australian Indigenous Languages have been subdivided into 3-digit categories.

Code length	Hierarchical level	Example
1-digit	Broad group	8 Australian Indigenous Languages
2-digit	Narrow group (2-digit)	81 Arnhem Land and Daly River Region Languages
3-digit	Narrow group (3-digit)	817 Kunwinjkuan
4-digit	Language	8171 Gundjeihmi

The three narrow groups which have been divided into 3-digit categories are:

- narrow group 81 Arnhem Land and Daly River Region Languages
- narrow group 82 Yolngu Matha and
- narrow group 86 Arandic

The ASCL code scheme is devised so that any future changes to the classification structure are easily accommodated. For example, when a language is added to ASCL it is allocated a previously unused code, and when a language is deleted from ASCL its code is retired and not used again.

Codes for residual categories

Not elsewhere classified (nec)

Any language which is not separately identified in the classification (because it does not meet the threshold for the minimum number of speakers) is included in the residual 'nec' category of the narrow group to which it belongs. NEC categories are easily identified as 4-digit codes ending with 99. Examples include: 1599 Scandinavian nec, 4299 Middle Eastern Semitic

Languages nec, and 6199 Burmese and Related Languages nec. ASCL contains 44 "nec" categories.

'Other' narrow group categories

Special codes are also reserved for residual categories at the narrow group level for separately identified languages which do not fit into any of the narrow groups within the broad group. They are termed 'Other' and are identified by the broad group code followed by '9'. ASCL contains seven such residual categories. Examples include: Narrow Group 39 Other Eastern European Languages (which includes Albanian and Romanian), and Narrow Group 49 Other Southwest and Central Asian Languages (which includes Armenian and Georgian). ASCL contains seven 'Other' categories.

Residual categories are part of the ASCL structure and should not be created or used to 'dump' responses which contain insufficient information to code to a separately identified category of the classification (see Supplementary Codes).

Supplementary codes

Supplementary codes are used to process inadequately described responses in statistical collections. There are two types of supplementary codes:

- 'Not further defined' ('nfd') codes are four-digit codes ending with either one, two or three zeros and are used to classify responses to questions about language which cannot be coded to the detailed four-digit language level of the classification but which clearly belong to a higher level category of the classification. For example, responses which do not relate directly to a detailed language category, but which are within the range of languages relating to a particular narrow group, are coded to that narrow group. Such responses are allocated a 'nfd' code consisting of the two digit code of the narrow group followed by 00. Examples of responses and their relevant nfd codes include: "China", which is coded to 7100 Chinese nfd, and "African", which is coded to 9200 African Languages nfd.

  Language responses which do not relate directly to a particular narrow group or language category, but are within the range of languages relating to a particular broad group, are coded to that broad group. These responses are allocated a 'nfd' code consisting of the one digit code of the broad group followed by the digits '000'. Examples of such responses and their relevant nfd codes include: "Indian", which is coded to 5000 Southern Asian Languages nfd, and "Aboriginal", which is coded to 8000 Australian Indigenous Languages nfd.

  'Not further defined' codes allow language responses which can be coded only at the broad or narrow group levels of the classification to be stored and processed within a collection coded at the four-digit level.

- 'Operational' codes are four-digit codes commencing with three zeros. These are used to code responses which contain minimal or vague information which cannot be coded to a language, a narrow group or even a broad group code. Examples of such responses and the relevant operational codes include: "many" or "multilingual" which are coded to 0000 Inadequately described, and "baby" and "not speaking yet" which are coded to 0001 Non-verbal so described.

  The code 0002 Not Stated is used when no response is given.

Index for coding responses

Why we use it

Responses provided in statistical and administrative collections are not always identical to the names used to describe the classification categories. Therefore, a coding index is required to link responses to the most appropriate code in the Australian Standard Classification of Languages (ASCL) in a process called "coding" (which can be undertaken by computer or manually). The ASCL coding index contains a comprehensive list of the most likely responses to questions relating to language and their correct classification codes. The coding index is used to code responses to questions such as 'First Language Spoken', 'Languages Spoken at Home', 'Main Language Spoken at Home' and 'Main Language Other Than English Spoken at Home'. The ASCL coding index may be requested by contacting standards@abs.gov.au (mailto:standards@abs.gov.au) .

How it was developed

The coding index was developed through literature research, consultation with stakeholders, and analysis of data from

responses obtained in ABS statistical collections such as the Census of Population and Housing.

As well as individual languages, a number of entries in the ASCL coding index cover dialects and regional language varieties not separately identified in ASCL. Therefore, in addition to its coding function, the numerical index can be used to clarify the nature, extent and varietal content of each language category.

Coding rules

When coding responses in statistical or administrative collections, the following rules apply:

- Responses which match exactly an entry in the coding index are assigned the code allocated to that index entry. For example, a response of "Deutsch" is coded to 1301 German, and "Cambodian" is coded to 6301 Khmer.

- Responses which relate directly to a language category are coded to that language category. Such instances include responses which are an exact match with the language category title except in terms of:
  - alternative spelling (e.g. responses of "Kaura", "Coorna" and "Koornawarra" are all coded to 8921 Kaurna)
  - spelling error (e.g. "Japanease" is coded to 7201 Japanese)
  - the use of abbreviations (e.g. "N.Z Maori" is coded to 9304 Maori (New Zealand))
  - the use of foreign or idiosyncratic words (e.g. "Nihongo" is coded to 7201 Japanese and "Aussie Slang" is coded to 1201 English)
  - the use of qualifying, modifying or extraneous words in addition to the fundamental or basic language description. For example, a response of "A little Japanese" or "Yes Japanese" is coded to 7201 Japanese and "South Korean" is coded to 7301 Korean.

- Responses which relate directly to a language category because they describe a variety, dialect or geographic variation of the language are coded to that language category (e.g. the responses "Swabian", "Viennese" and "Alsatian" are all coded to 1301 German).

- Responses containing more than one distinct language are coded to the first language stated (e.g. a response of "Polish and German" is coded to 3602 Polish). The exception to this rule is where it is possible to store more than one language code, in which case the code for each separate language is recorded.

- Responses which cannot be identified as relating to a separately identified language in the classification are assigned a residual category code or a supplementary code. For example "Chin" and "Chin Burma" are coded to 6100 Burmese and Related Languages nfd and "North Queensland Aboriginal" is coded to 8000 Australian Indigenous Languages nfd. Responses such as "Foreign", "Good Speech" and "Truth" cannot be linked to any language and are coded to 0000 Inadequately described.

A response should be coded to a residual category only when it is clear that it is a distinct language or dialect which cannot be placed in a precise language category. Responses which are not precise enough to be coded to any category should be assigned the appropriate supplementary code.

Using the classification

Editing specifications

It is important when validating input codes at editing stage, manipulating data, and deriving output items, that all valid codes are included in every specification. The full range of valid codes consists of all the codes in the classification structure plus all supplementary codes.

Coding, storage and presentation of data

Data should be collected, classified and stored at the language (four-digit) level of the classification to allow flexibility of statistical output and more detailed analysis. It also maintains information for future use and enables comparison with previous data using different classifications.

In some instances, concerns about confidentiality or standard errors may not permit the collection or output of data at the finer levels of the classification. The use of a standard classification enhances data comparability even though it may not always be possible to disseminate data at the most detailed level.

The hierarchical structure of the classification provides users the flexibility to output statistics at the level of the classification which best suits their particular purposes. Data can be presented at the broad group level, narrow group level, or the language level. Also, significant languages within a narrow group can be presented separately with the remaining languages of the narrow group aggregated.

Data downloads

> A coding index has been removed from the data cube. A coding index may be of use to anyone seeking to code responses to the Australian Standard Classification of Languages and may be requested by contacting standards@abs.gov.au (mailto:standards@abs.gov.au) .

Australian Standard Classification of Languages (ASCL), 2016
⤓ **Download XLS**
[250 KB]

Abbreviations

Show all

History of changes

Show all

Previous catalogue number

This release previously used catalogue number 1267.0